

REMARKS

The Office Action mailed on June 4, 2007 has been given careful consideration by applicant. Reconsideration of the application is requested in view of the amendments and comments herein. Claims 1, 16, 18, and 20 have been amended.

The Office Action

Claims 1, 8-13 and 15 are rejected under 35 U.S.C. §103(a) as being unpatentable over Kubota (US Patent No. 6,041,323) in view of Grefenstette (US Patent No. 6,396,951);

Claims 3-4, 6-7 and 16-21 are rejected under 35 U.S.C. §103(a) as being unpatentable over Kubota in view of Grefenstette and further in view of Gilfillan et al. (US PG Pub. No. 2002/0165856);

Claims 14 and 22 are rejected under 35 U.S.C. §103(a) as being unpatentable over Kubota in view of Grefenstette and further in view of Cofino et al. (US PG Pub. No. 2005/0187931).

First Obviousness Rejection

The Examiner has rejected claims 1, 8-13 and 15 under 35 U.S.C. §103(a) as being unpatentable over Kubota (US Patent No. 6,041,323) in view of Grefenstette (US Patent No. 6,396,951). This rejection should be withdrawn for at least the following reasons. Kubota and Grefenstette individually and in combination do not teach or suggest the subject invention as set forth in the subject claims.

Independent claim 1 includes all the limitations previously recited. In addition, a method is provided for identifying output documents that are similar to an input document. The input document that includes textual content is received and optical character recognition is performed on the textual content to identify text. The text and the textual content are analyzed to identify keywords, wherein a predefined number of keywords is identified from a first list of rated keywords extract from the input document. A list of best keywords is created for each keyword remaining in the first list of keywords, four steps are performed. 1) the keyword is identified in one or more domain specific dictionaries of words and phrases in which they are used; 2) combinations of keywords are identified on the list of keywords that satisfy the longest phrase; 3) the frequency of occurrence is

determined in the input document of the identified keywords and phrases identified in the one or more domain specific dictionaries; 4) and the linguistic frequency is of occurrence of the keywords and phrases is set to a predefined value. A list of best keywords is defined wherein the list of best keywords have rating greater than other keywords in the first list of keywords except for keywords belonging to a domain specific dictionary of words and having no measurable linguistic frequency. A query is formulated and performed utilizing a list of best keywords and a measure of similarity between the input document and each identified output document is computed. A second set of documents is defined from the first set of documents wherein a computed measure of similarity with the input document is greater than a predetermined threshold value. A matching document is utilized to identified similar documents wherein if one or more documents are related to a copyright registered document, than one or more documents is rights limited. Each document in the second set of documents is delivered to a predetermined output device. Kubota and Grefenstette individually and in combination do not teach or suggest the subject invention as set forth in the subject claims.

In particular, Kubota or Grefenstette do not teach or suggest one or more documents identified that are similar to an input document are determined to be related to a copyright registered document as recited in the subject claims. Kubota is concerned with a document search that may accommodate new words or phrases utilizing a request of a user for such search. A unique character string is extracted from an input document and a similarity search is performed by using the unique character string. Documents that are found are evaluated and arranged in order of evaluation. However, Kubota does not teach or suggest a determination as to whether one or more documents identified is a copyright registered document and further if such document is copyright registered the document is rights limited. Grefenstette does not make up for this deficiency.

In addition, Kubota does not teach or suggest defining a list of best keywords wherein the list of best keywords has a rating greater than other keywords in the first list of keywords except for keywords belonging to a domain specific dictionary of words and having no measurable linguistic frequency. The Examiner alleges that such limitation is taught by mere rearrangement of located documents in the order of evaluation. Further, the Examiner contends that character strings similar to a unique character strings are

performed via a search key dictionary. Since the search key dictionary has no errors in its list, the best keywords are less than the words contained in the search key dictionary. Utilizing a search key dictionary to evaluate similarity of character strings to a unique character string does not teach or suggest defining a list of best keywords belonging to a domain specific dictionary of words or having no measurable linguistic frequency. Instead Kubota does not utilize such factors in its evaluation and is silent regarding the elimination of defining best keywords relative to whether they belong in a domain specific dictionary of words and/or have no measurable linguistic frequency. Greffenstette does not make up for this deficiency.

Moreover, Kubota does not teach or suggest delivering each document identified as similar to an input document to a predetermined output device as recited in the subject claims. Instead Kubota contemplates storage of a program utilized to identify an input document input into a computer system wherein the computer system contains a comparison document searchably stored by a computer. Such storage medium may include a flopping disk CD ROM or storage device connected to a network. However, Kubota teaches storage of a program utilized to compare an input document to one or more despaired documents whereas the subject claims recite delivering a document to one or more output devices. Greffenstette does not make up for this deficiency.

For at least the aforementioned reasons, Kubota and Greffenstette individually and in combination do not teach or suggest the subject invention as recited in independent claim 1 (or claims 9-13, and 15, which depend therefrom). Accordingly, withdrawal of this rejection is respectfully requested.

Kubota such not teach or suggest repeating a search using the matching document to identify similar documents if the second set of documents includes a matching document but no similar documents, as recited in the subject claims. The Examiner cites col. 3, lines 63-66 of Kubota to teach this limitation. (A comparison document stored in the storage medium "[can be], in the case of multiple documents...a set of documents including the input document, or a set of document [sic] extracted by search or the like. The contents of document [sic] may be of a natural language or a program language."). This citation is improper since storing the input document as a comparison document does not teach or suggest a document matching the input document that is employed to identify similar

documents when no similar documents are originally located. This mechanism is in place to allow for a greater number of search results and is not contemplated or disclosed by Kubota.

For at least the aforementioned reasons, Kubota does not teach or suggest the subject invention as recited in independent claim 1 (or claims 9-13, and 15 which depend therefrom). Accordingly, withdrawal of this rejection is respectfully requested.

Second Obviousness Rejection

The examiner has rejected claims 3-4, 6-7 and 16-21 under 35 U.S.C. §103(a) as being unpatentable over Kubota in view of Grefenstette and further in view of Gilfillan et al. (US PG Pub. No. 2002/0165856). This rejection should be withdrawn for at least the following reasons. Kubota, Grefenstette and Gilfillan individually or in combination do not teach or suggest the subject invention as set forth in the subject claims.

Independent claim 16 (and similarly independent claims 18 and 20) recites limitations previously presented. For a method for computing ratings of keywords extracted from an input document. Determining if each keyword in the list of keywords exists in a domain specific dictionary of words by tokenizing the keywords at one or more predefined word boundaries. Frequency of occurrence is determined in the input document for each keyword in the list of keywords. Each keyword in the list of keywords a rating corresponding to its importance in the input document is computed that is a function of its frequency of occurrence in the input document and its frequency of occurrence in the collection of documents. This query is repeated until a predetermined number of results are obtained or the query is terminated. If one or more documents are identified in the collection of documents is a copy of a known copyright registered document, such document is rights limited. Each document in the collection of documents is delivered to a predetermined output device wherein the collection of documents is set forth in a list serialized in XML that contains for each document found: its location on a network, original representation, unformatted representation, service results, metadata, distance measurement, type of document found according to desired quality, and error status. Kubota, Grefenstette and Gilfillan individually or in combination do not teach or suggest the subject invention as set forth in the subject claims.

In particular, Kubota, Grefenstette and Gilfillan do not teach or suggest delivering identified similar documents to a predetermined output device wherein the set of documents is set forth in a list serialized in XML. Kubota, Grefenstette and Gilfillan do not each or contemplate the utilization of extensible markup language. Moreover, none of these references provide a list serialized in XML that contains for each document found its location on a network, original representation, unformatted representation, service results, metadata, distance measurement of document found according to desired quality, and error status. Accordingly, these references do not each or suggest the subject invention as set forth in the subject claims.

Moreover, none of these references teach or suggest creating a list of best keywords wherein each keyword remaining in a first list of keywords performs identifying the keyword in one or more domain specific dictionaries of words and phrases in which they are used; identifying combinations of keywords and the list of keywords that satisfy the longest phrase; determining the frequency of occurrence in the input document of the identified keywords and phrases identified in the one or more specific dictionaries; setting the linguistic frequency of occurrence of the keywords and phrases to a predefined value. None of the references teach or suggest identifying combinations of keywords and a list of keywords and a list of keywords that satisfy a longest phrase. Rather a unique character string is extracted from an input document and a similarity is searched utilizing such unique character string. There is no teaching and Kubota does not contemplate identifying combinations of keywords and a list of keywords that satisfy the longest phrase.

Neither Kubota nor Gilfillan teach or suggest repeating a search using the matching document to identify similar documents if the second set of documents includes a matching document but no similar documents, as recited in the subject claims. The Examiner cites col. 3, lines 63-66 of Kubota to teach this limitation. (A comparison document stored in the storage medium “[can be], in the case of multiple documents...a set of documents including the input document, or a set of document [sic] extracted by search or the like. The contents of document [sic] may be of a natural language or a program language.”). This citation is improper since storing the input document as a comparison document does not teach or suggest a document matching the input document that is employed to identify similar documents when no similar documents are originally located. This mechanism is in

place to allow for a greater number of search results and is not contemplated or disclosed by Kubota or Gilfillan.

For at least the aforementioned reasons, Kubota, Grefenstette and Gilfillan individually and in combination do not teach or suggest the subject invention as recited in independent claims 16, 18, or 20 (or claims 17, 19, and 21 which respectively depend therefrom). Moreover, claims 3-4 and 6-7 depend from independent claim 1 and Gilfillan does not make up for the aforementioned deficiencies of Kubota and Grefenstette regarding delivering each document identified to a predetermined output device. Accordingly, withdrawal of this rejection is respectfully requested.

Third Obviousness Rejection

The examiner has rejected claims 14 and 22 under 35 U.S.C. §103(a) as being unpatentable over Kubota in view of Grefenstette and further in view of Cofino et al. (US PG Pub. No. 2005/0187931). This rejection should be withdrawn for at least the following reasons. Claims 14 and 22 depend from independent claims 1 and 18 respectively, and Cofino et al. does not make up for the aforementioned deficiencies of Kubota, Grefenstette, or Gilfillan regarding delivering each identified document in the set of documents to a predetermined output device or there is a copy of a known copyright registered document. The one or more documents is rights limited. Thus, for at least the reasons discussed above with respect to claims 1 and 18, the combination of Kubota, Grefenstette and Cofino et al. do not teach or suggest the subject claims. Accordingly, the rejection of these claims should be withdrawn.

CONCLUSION

For the reasons detailed above, it is submitted all claims remaining in the application (Claims 1, 3-4, 6-7 and 9-22) are now in condition for allowance. The foregoing comments do not require unnecessary additional search or examination.


No additional fee is believed to be required for this Amendment. However, the undersigned attorney of record hereby authorizes the charging of any necessary fees, other than the issue fee, to Xerox Deposit Account No. 24-0037.


In the event the Examiner considers personal contact advantageous to the disposition of this case, he/she is hereby authorized to call Mark Svat, at Telephone Number (216) 861-5582.

Respectfully submitted,

FAY SHARPE LLP

11-5-07
Date


Mark Svat, Reg. No. 34,261
Kevin M. Dunn, Reg. No. 52,842
1100 Superior Avenue, Seventh Floor
Cleveland, OH 44114-2579
216-861-5582

<u>CERTIFICATE OF MAILING OR TRANSMISSION</u>	
I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being <input type="checkbox"/> deposited with the United States Postal Service "Express Mail" service under 37 CFR 1.10, addressed to: Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date indicated below. <input checked="" type="checkbox"/> transmitted to the USPTO by electronic transmission via EFS-Web on the date indicated below.	
Express Mail Label No.:	Signature: 
Date: <u>11-5-07</u>	Name: Elaine M. Checovich